Discrete and continuous univariate modelling

van Groenendaal, W.J.H.

Publication date:
1984

FEW

subfaculteit der econometrie

RESEARCH MEMORANDUM

[redacted]

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Postbus 90153 - 5000 LE Tilburg
Netherlands

[redacted]

FEW
141

Discrete and continuous univariate modelling

W.J.H. van Groenendaal

augustus 1983

Contents

0. Introduction.

This report describes the programming facility developed at the hybrid computer of the Technical University of Delft (TUD). The report is written by drs. W.J.H. van Groenendaal but the routines mentioned and the patching used are developed in close co-operation with ir. M. de Bruyn of the Hybrid Computer Centre of the Technical University of Delft.

The report consists out of four sections. Section one contains a short introduction on hybrid programming and the hybrid computer including a definition of the components available for calculation. These components are used to define a patching for two relations which are used to obtain least squares estimates for models for which they fit. It is further analysed in which way the use of the hybrid computer can influence the estimation results.

In section two a number of possible discrete and continuous models are given which can be estimated using the patching outlined in section one.

In section three some applications are given and section four contains a number of concluding remarks.

I. Hybrid Optimization

I.1. The computer configuration used.

A traditional digital computer reduces a problem into a sequence of bytes and this implies that a digital computer is a sequentially working computer. In more advanced digital computers the first elements of parallel computation are introduced which means that several calculations can be performed at the same time. This seems revolutionary from a users point of view but it is not. It is not because one of the first computers, the analog computer, is also a parallel working computer.

On an universal*) analog computer a problem is converted into an electrical network which is called the patching of the problem. Within this network quantities are represented by voltages which can vary continuously. In this way a problem like a simultaneous model can be simulated at once without the necessity of special simulation algorithems. Because of this property an analog computer is a parallel processor.

If one combines a digital and an analog computer by means of an interface, which enables communication between the two computers, one gets a hybrid computer. On a hybrid computer one can use the features of both computers. It is also possible to combine several analog computers into one larger analog computer which is called slaving. The capacity of this combination is equal to the sum of the capacities of the analog computers combined.

The analog computer used in this study is the AD4 ≠ 1 of the TUD which is combined with a PDP 11/45 into a hybrid computer.

*) Universal because there also exist special purpose analog computers.

I.2. A short Introduction into Hybrid Programming.

Because of the differences between the two types of computers which are combined into a hybrid computer, programming has to be performed carefully. It is beyond the scope of this report to explain in detail the ins and outs of hybrid programming and only those aspects which are necessary for the understanding of the next paragraphs are reported here.

Scaling

The first aspect of interest is the necessity of scaling. Because the quantities of a problem implemented on an analog computer are represented by voltages they are limited to a certain range. ([-100 volt, 100 volt])
This range is refered to as the machine range and is set equal to [-1,1]. Also the variables one wants to simulate or calculate are within a certain range which in general is not equal to the machine range. Therefore it will be necessary to transform the model variables into machine variables. How this can be done will be illustrated by the following example.

Suppose one wants to implement the linear model

$$(I.2.1)\quad y(t) = \alpha_0 + \alpha_1 x_1(t) + \alpha_2 x_2(t), \qquad t \in [t_0, t_1]$$

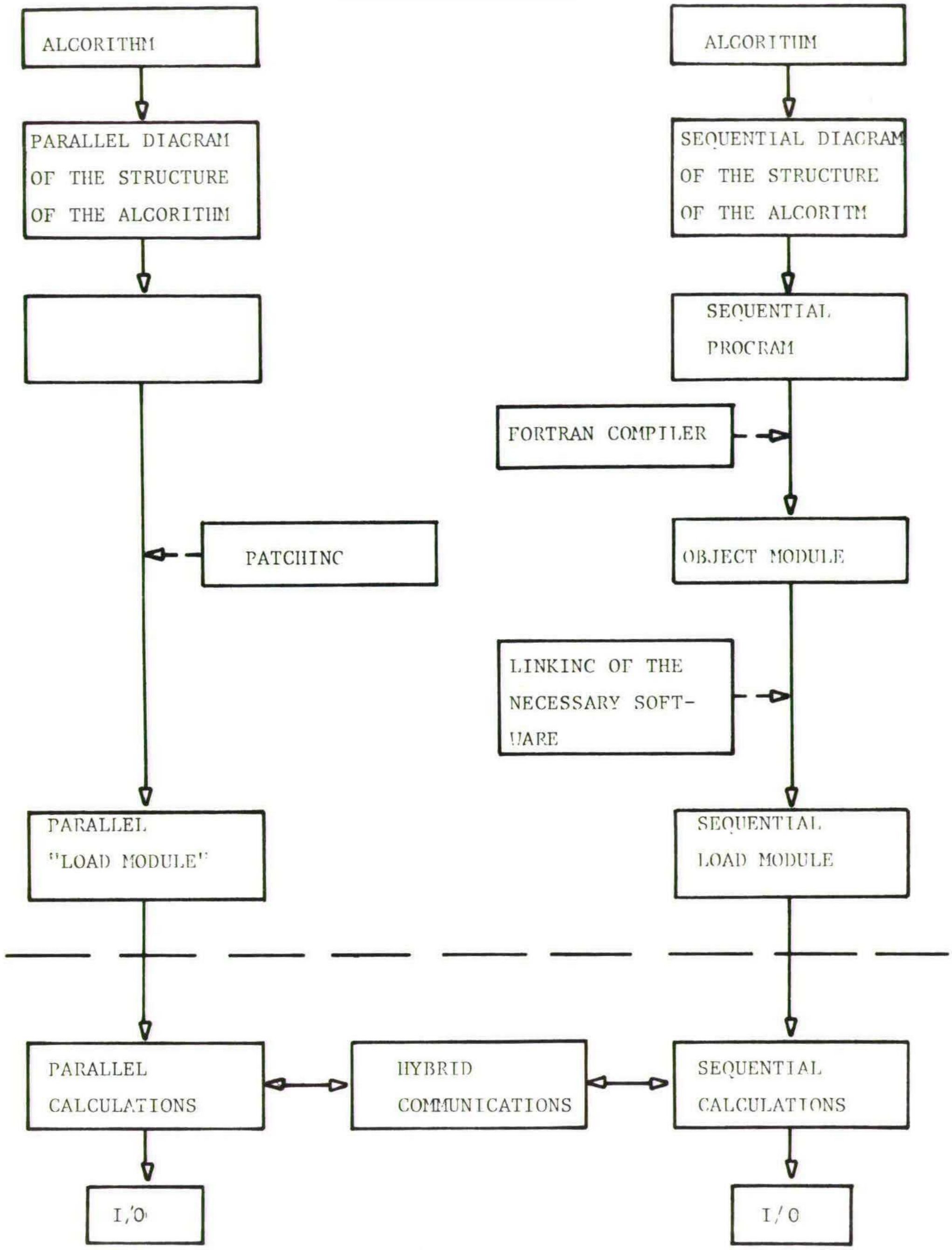
The ranges of the model variables $y(t)$, $x_1(t)$ and $x_2(t)$ on $[t_0, t_1]$ are known*) so one can calculate $\hat{y}$, $\hat{x}_1$ and $\hat{x}_2$ which are the maximum values of respectively $y(t)$, $x_1(t)$ and $x_2(t)$ on $[t_0, t_1]$. The transformation into the machine range of the model (1) can be performed by rewritting model (I.2.1) into

$$(I.2.2)\quad y'(t)\hat{y} = \alpha_0 + \alpha_1 \hat{x}_1 x_1'(t) + \alpha_2 \hat{x}_2 x_2'(t), \; t \in [t_0, t_1] .$$

*) Otherwise one has to deduce an upper- and lowerbound.

Fig. I.2.1

Hybrid programming*)

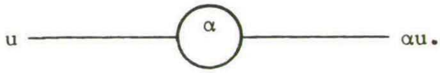


*) Source: Introduction in hybrid programming by M. de Bruyn.

I.3. Summary of the components.

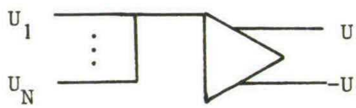
In this section the components with which one can design the physical representation of a model will be shortly introduced.

(i) Coefficient units. These units perform multiplication with a constant end are symbolized in the design of a patching by



The range of coefficients is [-1.6384, 1.6384) or [0,3.2768).

(ii) Summers. A summer makes a linear combination $U = \sum_{i=1}^{N} U_i$ and its design symbol is



The summer has two possible output signals +u and -u. The range is [-1,1].

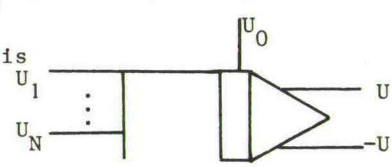
(iii) Integrators. An integrator integrates the differential equation

$$\frac{dU(t)}{dt} = \sum_{i=1}^{N} U_i(t), \; U(t_0) = U_0$$

which leads to

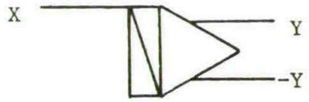
$$U(t) = U_0 + \sum_{i=1}^{N} \int_{t_0}^{t_1} U_i(t)dt, \; U_0 = U(t_0).$$

The symbol used in designs is



The range of an integrator is also [-1,1] and every integrator can also be used as a summer.

(iv) Track-store.

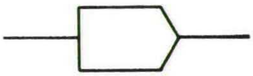


A track store has three possible outputs

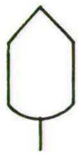
track: $y = \pm x$
store: y is constant from the moment it is store made
IC : y = IC.

It is clear that the track-store unit opens the possibility for storing the value of a variable for one or more time units within a patching. This can be used when simulating difference equations totally on the analog computer of for the simulation of mixed difference-differential equations.

(v) DAC's. The digital to analog converter (DAC) makes it possible to convert a digitally represented figure into an electrical signal. The symbol is



(vi) DSP. The output of a summer/integrator can be displayed on the memoscope by using the display facility. The symbol is



I.4. The Application of the Data.

There are two aspects of interest in the application of the data. One aspect is the technical aspect of how the data are transformed into an electrical signal of a certain voltage and the second aspect is how the data are made continuous.

The first aspect is only of technical interest. The data are applied to the interface where the digitally represented figure is converted into an electrical signal by means of the digital to analog converter (DAC).

The second aspect is of more interest. Here we will only treat the technical aspect. The data available are always discrete end because the parallel processor is working in continuous machine time one has to choose some form of continuous data. This is done by linear interpolation. The discrete values $z(\tau)$ and $z(\tau-1)$ of a variable $z(t)$ are applied to two different DAC's. Using a convex combination $(1-h(\theta))z(\tau) + h(\theta)z(\tau-1)$ with $h(\theta)$ a linear function which is zero if $\theta = \tau$ and one if $\theta = \tau-1$ one gets the desired result. The function $h(\theta)$ is called the saw-tooth generator.

From what follows in the next sections it will become clear that the fact that the computer is working in continuous time does not imply that the models used have to be continuous too. If, like in our case, optimization is performed on sampled data also discrete interpretations are possible.

The way the data are interpolated does also have no influence on the optimization result. The same result can by achieved if the data are applied as piece-wise constant functions. In both cases it is only necessary that the sampling takes place at those points in machine time that correspond to τ and $\tau-1$.

I.5. Definition of the patching.

With the use of the components defined in section I.3 a patching was defined for optimization of the forms

$$(I.5.1)\quad y(t) = \alpha_1^T v(t) + \alpha_2^T z_\lambda(t) + \alpha_0 + \varepsilon(t)$$

and

$$(I.5.2)\quad \frac{dy(t)}{dt} = \gamma y(t) + \alpha_1^T v(t) + \alpha_2^T z_\lambda(t) + \alpha_0 + \varepsilon(t)$$

with $z_\lambda(t)$ the solution of

$$\frac{dz_{\lambda_i}(t)}{dt} = \lambda_i(z(t) - z_{\lambda_i}(t));\ i = 1,\ldots,6.$$

This patching is given in figure (I.5.1). The restrictions are, $\lambda_i \in R^+$ α_0, $\gamma \in R$, $\alpha_1 \in R^4$ and $\alpha_2 \in R^6$.

From the patching we see that not all variables are applied through a pair of DAC's. This is because the number of DAC's is limited to 16. Sylwestrowicz concluded from this that only eight variables could be converted into a continuous signal (Sylwestrowicz (1979) pag. 27). However a solution for this problem is found in the use of coefficient units to simulate a DAC. In this way DAC's are no longer a restriction on the number of variables which can be transferred to the analog computer.

The patching (fig. I.5.1.) contains the possibility of display for all variables applied using display 1 to 10. On display 0 one gets the result for the endogenous variable.
In case of model (I.5.1) coefficient unit 250 is equal to zero and 252 equal to one. In case of model (I.5.2) the opposite is set. On displayline 12, the data are available. Because the user has the possibility to display several variables at once the fit on displayline 0 and the data can be compared. Within the patching also the difference, the squared difference and the integral of the squared difference are calculated. This gives the opportunity to minimize

$$S(\theta) = \int_0^T (y(t) - \hat{y}(\theta;t))^2 dt$$

with $\tilde{y}(t)$ the linear interpolation of the data, instead of

$$S(\theta) = \sum_{i=1}^{N} (y(i\delta) - \hat{y}(\theta;i\delta))^2.$$

with δ the sampling period.

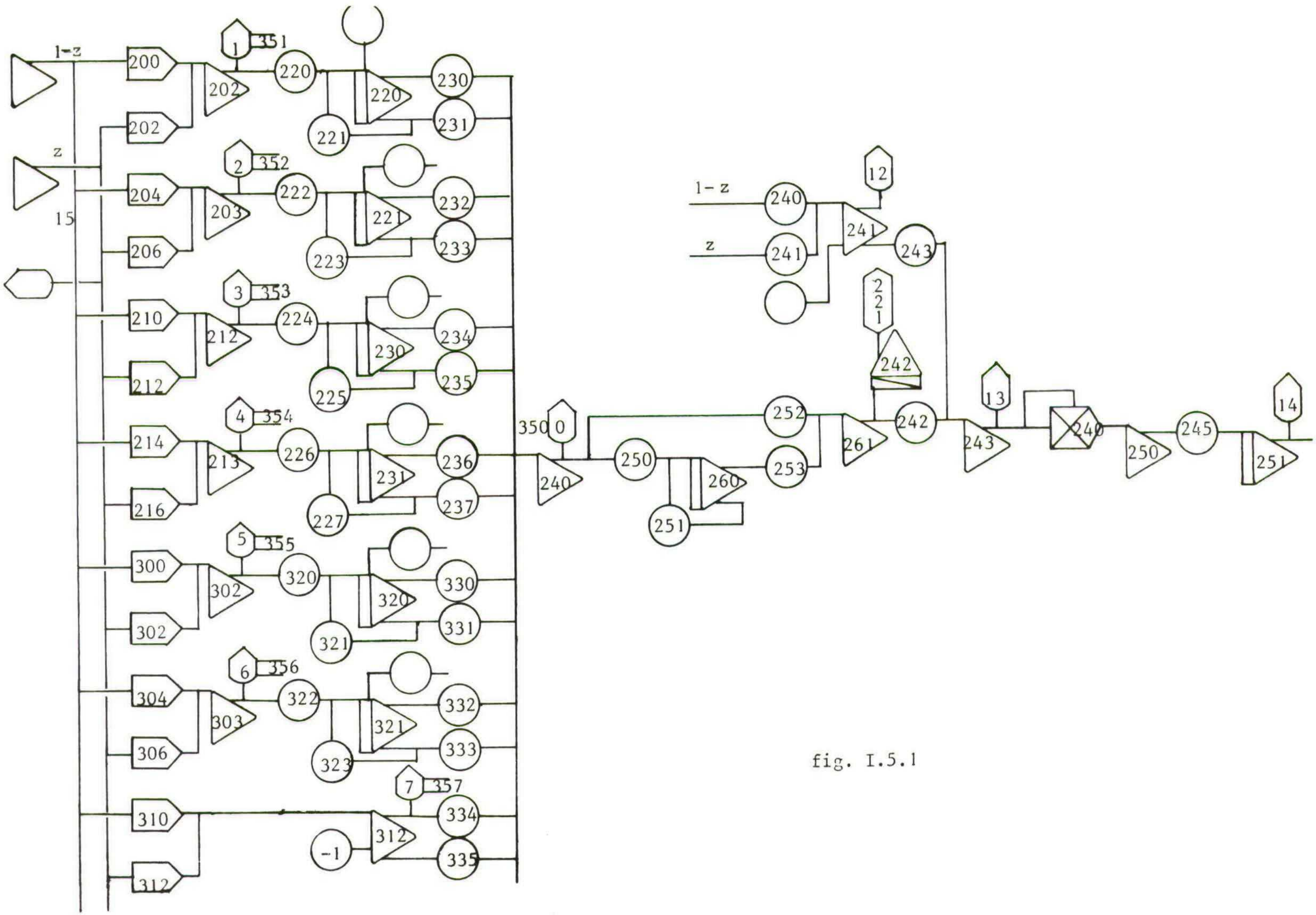
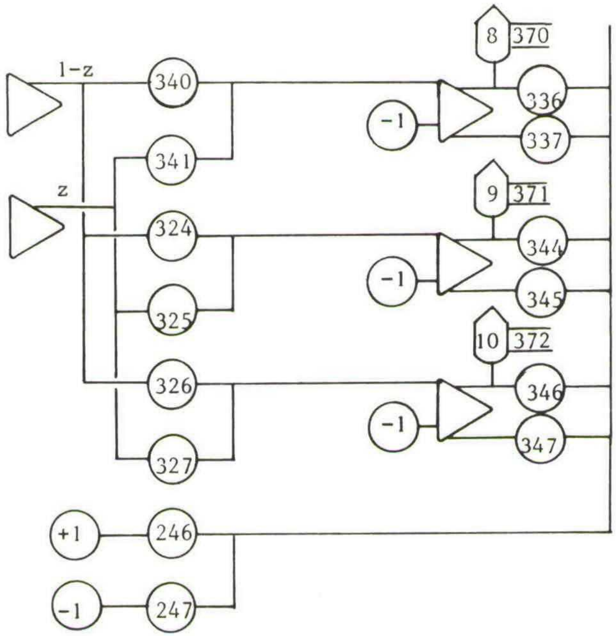

1-z
z
15
200
202
204
206
210
212
214
216
300
302
304
306
310
312
202
203
212
213
302
303
1
2
3
4
5
6
7
351
352
353
354
355
356
357
220
222
224
226
320
322
221
223
225
227
321
323
220
221
230
231
320
321
312
-1
230
231
232
233
234
235
236
237
330
331
332
333
334
335
350
0
240
250
251
260
252
253
261
1- z
z
240
241
241
243
12
2
2
1
242
242
243
13
240
250
245
251
14

fig. I.5.1


1-z
z
340
341
324
325
326
327
-1
-1
-1
8
370
9
371
10
372
336
337
344
345
346
347
+1
-1
246
247

I.6. Parameter Estimation.

Our aim is to determine "values of parameters that govern the dynamic and/or non-linear behaviour of the process under study, assuming that the structure of the process is well known." (Eykhoff [1974], pag. 8) In this report two processes are under study

$$(\text{I.6.1a})\quad y(t) = \beta_0 + \beta_1^T x_\lambda(t) + \beta_2^T x(t) + \varepsilon(t) \;^{*)}$$

with

$$(\text{I.6.1b})\quad \frac{dx_{i\lambda_i}(t)}{dt} = \lambda_i(x_i(t) - x_{i\lambda_i}(t));\; i = 1,\dots,p$$

and

$$(\text{I.6.2a})\quad \frac{dy(t)}{dt} = \alpha y(t) + \beta_0 + \beta_1^T x_\lambda(t) + \beta_2^T x(t) + \varepsilon(t)$$

with

$$(\text{I.6.2b})\quad \frac{dx_{i\lambda_i}(t)}{dt} = \mu_i(x_i(t) - x_{i\lambda_i}(t));\; i = 1,\dots,p$$

$\beta_0 \in R$; $\alpha \in R^-$; $\beta_1 = (\beta_{11},\dots,\beta_{1p}) \in R^p$, $\beta_2 = (\beta_{21},\dots,\beta_{2q}) \in R^q$
$x_\lambda(t) = (x_{1\lambda_1}(t),\dots,x_{p\lambda_p}(t)) \in R^p$, $x(t) = (x_{p+1}(t),\dots,x_{p+q}(t)) \in R^q$.

In both cases the least squares criterion was chosen to formulate the identification problem. In first instance it is regarded as a pure optimization problem. Assumptions with respect to the error process $\varepsilon(t)$ transforms the optimization problem into an estimation problem. In a number of cases the estimates have desirable properties. The criterion function is

$$(\text{I.6.3})\quad S(\theta) = (y'^{(t)} - \hat{y})^T(y'^{(\theta)} - \hat{y})$$

with $\theta = (\beta_0, \beta_1, \lambda, \beta_2) \in R^{2p+q+1}$ or $\theta = (\alpha, \beta_0, \beta_1, \lambda, \beta_2) \in R^{2p+q+2}$ the vector of parameters. $y, \hat{y} \in R^N$, y the vector of observations and $\hat{y}$

*) For the moment possible dead-time delays are neglected.

the vector of sampled simulation results obtained from the model

$$\text{(I.6.4a)}\quad \hat{y}(t) = \hat{\beta}_0 + \hat{\beta}_1 x_{\hat{\lambda}}(t) + \hat{\beta}_2 x(t)$$

$$\text{(I.6.4b)}\quad \frac{dx_{i\hat{\lambda}_i}(t)}{dt} = \hat{\lambda}_i(x_i(t) - x_{i\hat{\lambda}_i}(t));\ i = 1,\ldots,p$$

or from the model

$$\text{(I.6.5a)}\quad \frac{d\hat{y}(t)}{dt} = \hat{\alpha}\hat{y}(t) + \hat{\beta}_0 + \hat{\beta}_1 x_{\hat{\lambda}}(t) + \hat{\beta}_2 x(t)$$

$$\text{(I.6.5b)}\quad \frac{dx_{i\hat{\lambda}_i}(t)}{dt} = \hat{\lambda}_i(x_i(t) - x_{i\hat{\lambda}_i}(t));\ i = 1,\ldots,p$$

A problem is that even when the λ_i's are known we don't know $x_{i\lambda_i}(t)$ because $x_i(t)$ is only known over a limited period. It is assumed that $x_{i\hat{\lambda}_i}(t_0) = 0$ with t_0 outside the optimization period. Let t_1 be the starting point of the optimization period. If the $\hat{\lambda}_i$'s are "small" compared to (t_1-t_0) the effect of this initialization can be neglected. In practice two or three data periods appeared to be enough.

With the assumption $x_{i\lambda_i}(t_0) = 0$ it is now possible to obtain optimization results with respect to the vector of parameters. It is immediate that $S(\theta)$ is minimized if $(\beta_0, \beta_1, \lambda, \beta_2) = (\hat{\beta}_0, \hat{\beta}_1, \hat{\lambda}, \hat{\beta}_2)$ or $(\alpha, \beta_0, \beta_1, \lambda, \beta_2) = (\hat{\beta}_0, \hat{\beta}_1, \hat{\lambda}, \hat{\beta}_2))$. *)

There is however one more problem. Because of the nature of the analog computer the sampled results are corrupted by random noise. The effects of this noise on the optimization results are roughly analyzed in the next section.

*) For a description of the optimization procedure see Sylwestrowicz 1979.

I.7. Optimization in a hybrid environment.

Because of the nature of the analog computer the sampled results are corrupted with random noise. This noise is a result of the physical design of the analog computer and will influence the result of the minimization of the object function $S(\theta)$. The nature of this influence is subject of this paragraph. In first instance the simple model

$$(I.7.1a)\quad y(t) = \alpha_0 + \alpha_1 z_\lambda(t) + \alpha_2 v(t) + \varepsilon(t)$$

$$(I.7.1b)\quad \frac{dz_\lambda(t)}{dt}\ \lambda(z(t) - z_\lambda(t))$$

will be analyzed. The least squares result is

$$(I.7.2a)\quad \hat{y}(t) = \hat{\alpha}_0 + \hat{\alpha}_1 z_{\hat{\lambda}}(t) + \hat{\alpha}_2 v(t)$$

$$(I.7.2b)\quad \frac{dz_{\hat{\lambda}}(t)}{dt} = \hat{\lambda}(z(t) - z\hat{\lambda}(t))$$

The analog implementation of this form can be represented by

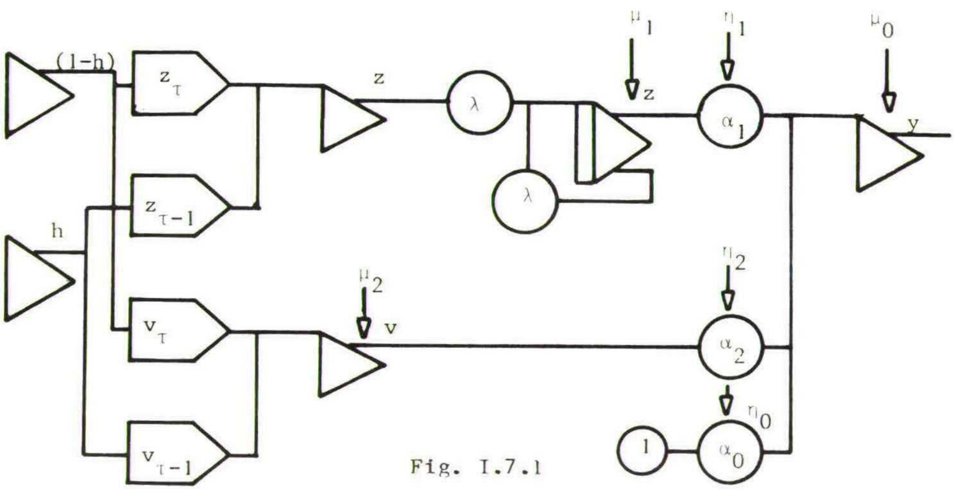


Fig. I.7.1

For a description of the symbols used see section I.3.
Because of the physical noise not the value z(t) but a stochastic variable $\bar{z}(t)$ is the result of the digital to analog conversion. The same is true for v(t).

Also the coefficient units are corrupted with noise. What we are interested in at this point is in what way the noise influences the optimization results. It is therefore assumed that it can be represented as

$$w(t) = \bar{y}(t) + \mu_0(t) \text{ *)}$$
$$\bar{z}_{\bar{\lambda}}(t) = z_{\hat{\lambda}}(t) + \mu_1(t)$$
$$\bar{v}(t) = v(t) + \mu_2(t)$$
$$\bar{y}(t) = \bar{z}_{\bar{\lambda}}(t) + \bar{v}(t)$$

With respect to the noise variables $\mu_1(t)$ and $\mu_2(t)$ it is assumed that they are ergodic, i.e. the ensemble and time moments are the same. It is also assumed that

$$E\{\mu_i(t)\} = 0 \text{ and } E\{\mu_i(t), \mu_j(t)\} = \delta_{ij}\sigma^2_{\mu_i} \qquad i,j = 0,1,2.$$

where δ_{ij} is the kronecker delta. The coefficients can be represented by

$$\bar{\alpha}_0(t) = \hat{\alpha}_0 + \eta_0(t)$$
$$\bar{\alpha}_1(t) = \hat{\alpha}_1 + \eta_1(t)$$
$$\bar{\alpha}_2(t) = \hat{\alpha}_2 + \eta_2(t)$$

and with respect to the noise process the same assumptions are made, the processes are ergodic and independent.

The implemented form can be described by

$$\text{(I.7.3a)} \quad \bar{y}(t) = \bar{\alpha}_0(t) + \bar{\alpha}_1(t)\bar{z}_{\bar{\lambda}}(t) + \bar{\alpha}_2(t)\bar{v}(t)$$

$$\text{(I.7.3b)} \quad \frac{d\bar{z}_{\bar{\lambda}}(t)}{dt} = \bar{\mu}(\bar{z}(t) - \bar{z}_{\bar{\lambda}}(t))$$

and introduction of the assumptions in (I.7.3a) leads to

$$\bar{y}(t)=(\hat{\alpha}_0+\eta_0(t))+(\hat{\alpha}_1+\eta_1(t))(z_{\hat{\lambda}}(t)+\mu_1(t))+(\hat{\alpha}_2+\eta_2(t))(v(t)+\mu_2(t))-\mu_0(t)$$

*) Apart from the noise of the preceeding summers etc. an extra complication here is the noise of the parameter which cannot be distinguished from the noise of the integrator.

$$\bar{y}(t)=\hat{\alpha}_0+\hat{\alpha}_1\hat{z}_\lambda(t)+\hat{\alpha}_2v(t)+\eta_0(t)+\eta_1(t)z_\lambda^{\hat{}}(t)+\eta_2(t)v(t)+\hat{\alpha}_1\mu_1(t)+\hat{\alpha}_2\mu_2(t)$$

$$+\eta_1(t)\mu_1(t)+\eta_2(t)\mu_2(t)-\mu_0(t)$$

$$\bar{y}(t)=\hat{y}(t)+\eta_0(t)+\eta_1(t)z_\lambda^{\hat{}}(t)+\eta_2(t)v(r)+\hat{\alpha}_1\mu_1(t)+\hat{\alpha}_2\mu_2(t)+\eta_1(t)\mu_1(t)+$$

$$\eta_2(t)\mu_2(t)-\mu_0(t)$$

From this it is clear that the sampled values are corrupted with noise and with its use the influence on the criterion function can be analysed. Let t_1, $t_2,\ldots,t_N$ be the sample points. Introduction of the sequence $\bar{y}(t_1),\ldots,\bar{y}(t_N)$ in criterion function $S(\theta)$ instead of ϑ leads to

$$\text{(I.7.5)}\quad \sum_{i=1}^{N}\Big[\{(\alpha_0-\hat{\alpha}_0)+(\alpha_1-\hat{\alpha}_1)z_\lambda(t_i)+\hat{\alpha}_1(z_\lambda(t_i)-z_\lambda(t_i))+(\alpha_2-\hat{\alpha}_2)v(t_i)\}^2+$$

$$\{\varepsilon(t_i)+\mu_0(t_i)+\eta_0(t_i)+\eta_1(t_i)z_\lambda(t_i)+\eta_2(t_i)v(t_i)+\hat{\alpha}_1\mu_1(t_i)+\hat{\alpha}_2\mu_2(t_i)$$

$$+\eta_1(t_i)\mu_1(t_i)+\eta_2(t_i)\mu_2(t_i)\}^2-2\{\alpha_0-\hat{\alpha}_0)+(\alpha_1-\hat{\alpha}_1)z_\lambda(t_i)+\alpha_1(z_\lambda(t_i)-$$

$$z_\lambda^{\hat{}}(t_i))(\alpha_2-\hat{\alpha}_2)v(t_i)\}\{\varepsilon(t_i)-\mu_0(t_i)+(\eta_0(t_i)+\eta_1(t_i)z_\lambda^{\hat{}}(t_i)$$

$$+\eta_2(t_i)v(t_i))+(\hat{\alpha}_1\mu_1(t_i)+\hat{\alpha}_2\mu_2(t_i))+(\eta_1(t_i)\mu_1(t_i)+\eta_2(t_i)\mu_2(t_i))\}$$

If it is assumed that $E\{\varepsilon(t_i)\mu_h(t_j)\}=0$; $h=0,1,2$; $E\{\varepsilon(t_i)\eta_k(t_j)\}=0$, $k=0,1,2$ and $E\{u_h(t_i)\eta_k(t_j)\}=0$, $h=0,1,2$, $k=0,1,2$; $i,j=1,\ldots,N$ and expectations are taken one gets

$$\text{(I.7.6)}\quad E\{\overline{S(\theta)}\}=\sum_{i=1}^{N}\Big[\{(\alpha_o-\hat{\alpha}_0)+(\alpha_1-\hat{\alpha}_1)z_\lambda(t_i)+\hat{\alpha}_1(z_\lambda(t_i)-z_\lambda^{\hat{}}((t_i))$$

$$(\alpha_2-\hat{\alpha}_2)v(t_i)\}^2+\sigma_\varepsilon^2+\sigma_\mu^2+\sigma_{\eta_0}^2+\sigma_{\eta_1}^2\,z_\lambda^{\hat{}}(t_i)^2+\sigma_{\eta_2}^2\,v(t_i)^2+\hat{\alpha}_1^2\sigma_{\mu_1}^2+\hat{\alpha}_2^2\sigma_{\mu_2}^2$$

Formula (I.7.6) shows that we may expect a positive deviation of the hybrid sum of squares compared with the digital sum of squares which would be obtained on a digital computer. This sum of squares consists out of the first two terms of (I.7.6) i.e.

$$E\{S(\theta)\} = \sum_{i=1}^{N} [\{(\alpha_0-\hat{\alpha}_0)+(\alpha_1-\hat{\alpha}_1)z_\lambda(t_i)+\hat{\alpha}_1(z_\lambda^\wedge(t_i)-z_\lambda^\wedge(t_i))+$$

$$(\alpha_2-\hat{\alpha}_2)v(t_i)\}^2 + \sigma_\varepsilon^2 \; .$$

A difficulty is that the positive bias depends on the parameters. However from the design of the analog computer we know that $\sigma_{\eta_i} < 0.003$ and $\sigma_{\mu_i} < 0.003$, $i = 0,1,2$.
Because the data are in the range [-1,1] and the parameters in the range [-1,5, 1,5] the positive bias is dominated by

$$\text{(I.7.7)} \quad \sum_{i=1}^{N} \sigma_{\mu_0}^2 + \sigma_{\eta_0}^2 + \sigma_{\eta_1}^2 z_\mu^\wedge(t_i)^2 + \sigma_{\eta_2}^2 v(t_i)^2 + \hat{\alpha}_1^2\sigma_{\mu_1}^2 + \alpha_2^2\sigma_{\mu_2}^2$$

$$< (0.003)^2*N*(1+1+1+1+2.25+2.25)$$

$$< (0.003)^2*N*8.5 = 76,5*N*10^{-6}$$

In general the positive bias will be dominated by

$$(0.003)^2*N*(2+3.25(p+q))$$

A trivial lowerbound is zero.

Generalization of model (I.7.1) is easily established

$$\text{(I.7.8)} \quad \begin{aligned} y(t) &= \alpha_1^T v(t) + \alpha_2^T z_\lambda(t) + \varepsilon(t) \\ \frac{dz_{\lambda_i}(t)}{dt} &= \lambda_i(z_i(t) - z_{\lambda_i}(t)) \qquad i = 1,\dots,p \end{aligned}$$

$y(t) \in R$, $\alpha_1^T = (\alpha_{10}, \alpha_{11},\dots,\alpha_{1q}) \in R^{q+1}$, $v(t) = (1,v,(t),\dots,v_q(t)) \in R^{q+1}$, $\alpha_2^T = (\alpha_{21},\dots,\alpha_{2p}) \in R^p$; $z_\lambda(t) = (z_{\lambda_1}(t),\dots,z_{\lambda_p}(t)) \in R^p$, $\varepsilon(t) \in R$.

The implemented model is

$$\bar{w}(t) = \bar{\alpha}_1(t)^T\bar{v}(t) + \bar{\alpha}_2(t)^T\bar{z}_{\bar{\lambda}}(t)$$

$$\text{(I.7.9)}\quad \frac{d\bar{z}_{\bar{\lambda}_i}(t)}{dt} = \bar{\lambda}_i(\bar{z}_i(t) - \bar{z}_{\bar{\lambda}_i}(t)) \qquad i = 1,\ldots,p$$

It is assumed that

$$\bar{\alpha}_{1i}(t) = \hat{\alpha}_{1i} + \eta_{1i}(t),\ i = 0,\ldots,q$$

$$\bar{\alpha}_{2i}(t) = \hat{\alpha}_{2i} + \eta_{2i}(t),\ i = 1,\ldots,p$$

where the η-processes are ergodic and independent and

$$\bar{w}(t) = \bar{y}(t) + \mu_0(t)$$
$$\bar{v}_i(t) = v_i(t) + \mu_i(t);\ i = 1,\ldots,q$$
$$\bar{z}_{\bar{\lambda}_i}(t) = z_{\hat{\lambda}_i}(t) + \mu_{2i}(t):\ i = 1,\ldots,p$$

with the same assumptions for the μ-processes as for the η-processes except for μ_{10} which is zero. $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda})$ represents the desired L.S-result. Introduction of the assumptions in (I.7.9) leads to

$$\bar{y}(t) = (\hat{\alpha}_1+\eta_1(t))^T(v(t)+\mu_1(t)) + (\hat{\alpha}_2+\eta_2(t))^T(z_{\hat{\lambda}}(t) + \mu_2(t)) + \mu_0$$

(I.7.10)

$$\bar{y}(t) = \hat{\alpha}_1^T v(t)+\hat{\alpha}_2^T z_{\hat{\lambda}}(t)+\eta_1(t)^T v(t)+\eta_2(t)^T z_{\hat{\lambda}}(t)+\hat{\alpha}_1^T\mu_1(t)+\hat{\alpha}_2^T\mu_2(t)$$

$$+ \eta_1(t)^T\mu_1(t) + \eta_2(t)^T\mu_2(t) + \mu_0$$

Introduction of the sampled output $\bar{y}(t_i)$, $i = 1,\ldots,N$ instead of $\hat{y}(t_i)$ in the criterion function leads to

$$\text{(I.7.11)}\quad \overline{S(\theta)} = \{\sum_{i=1}^{N} (\alpha_1-\hat{\alpha}_1,\ \alpha_2-\hat{\alpha}_2)^T\binom{v}{z_\lambda}_i+\hat{\alpha}_2(z_\lambda-z_{\hat{\lambda}})_i+\varepsilon_i-\mu_{0_i}-$$

$$(\eta_1,\eta_2)_i^T\binom{v}{z_{\hat{\lambda}}}_i-(\hat{\alpha}_1,\hat{\alpha}_2)^T\binom{\mu_1}{\mu_2}_i-(\eta_1,\eta_2)_i^T\binom{\mu_1}{\mu_2}_i\}^2$$

The time index t_i is omited for notational convenience.

With the same additional assumption with respect to the noise processes as were made in the two variable case (independency) the bias is also a straight forward generalization.

$$E\{\overline{S(\theta)}\} = \Sigma[\{(\alpha_1-\hat{\alpha}_1,\ \alpha_2-\hat{\alpha}_2)^T\binom{v}{z_\lambda} + \hat{\alpha}_2(z_\lambda - z_{\hat{\lambda}})_i\}^2$$

$$+ \sigma_\varepsilon^2 + \sigma_{\eta_0}^2 + \sigma_{\eta_1}^{2^T} v_i^2 + \sigma_{\eta_2}^{2^T} \hat{z}_\lambda^2 + \hat{\alpha}_1^{2^T} \sigma_{\eta_1}^{2^T} + \hat{\alpha}_2^{2^T} \sigma_{\eta_2}^2$$

where $\sigma_{\eta_1}^2 = (\sigma_{\eta_{01}}^2, \sigma_{\eta_{02}}^2, \ldots, \sigma_{\eta_{1q}}^2)$ etc.

In our application the largest value for (p+q) is six and N = 30 so the positive bias is dominated by

$$(0.003)^2 * 30 * (2+3.25*6) < 12*10^{-4}$$

which is negligible.

Another model is the differential equation model

$$\frac{dy(t)}{dt} = \beta\ y(t) + \alpha_0 + \alpha_1 z_\lambda(t) + \alpha_2 v(t) + \varepsilon(t)$$

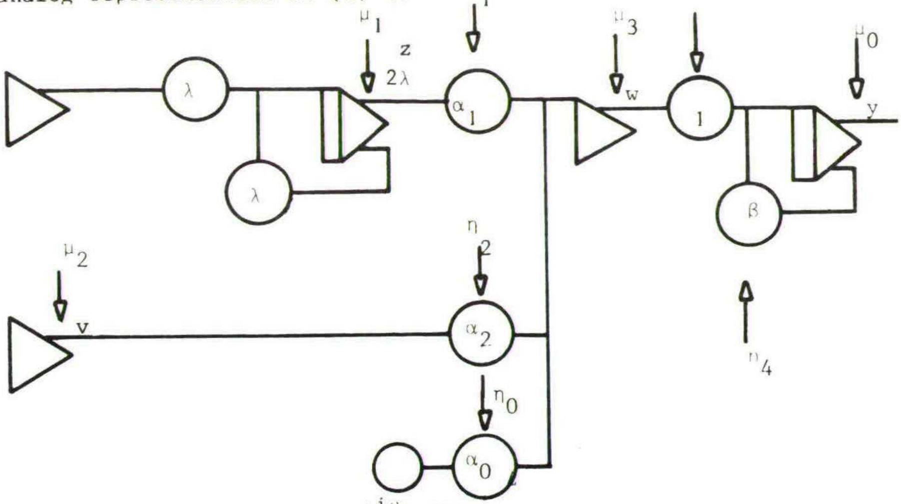
$$\frac{dz_\lambda(t)}{dt} = \mu(z(t) + z_\lambda(t))$$

The least squares result is represented by

$$\frac{d\hat{y}(t)}{dt} = \hat{\beta}\ \hat{y}(t) + \hat{\alpha}_0 + \hat{\alpha}_1 z_{\hat{\lambda}}(t) + \hat{\alpha}_2 v(t)$$

$$\frac{dz_{\hat{\lambda}}(t)}{dt} = \hat{\lambda}\ (z(t) - z_{\hat{\lambda}}(t))$$

The analog representation of (2) is



The interpretation of the variables is the same as for fig. (I.7.2). The noise structure is slighly different.
Apart from the processes distinguished in the preceeding section it is necessary to add $(1+\eta_3(t))$, $(\hat{\beta}+\eta_4(t))$ and $\bar{\bar{y}}(t) = \bar{y}(t)+\mu_3(t)$.

It is difficult to analyse this problem in the same way as the former, but we have seen that the error on $\bar{y}(t)$ was only small although we used the maximum error per component as an upperbound. From this we may expect that the total error due to machine noise in $\bar{\bar{y}}(t)$ will be small although it will be larger than in the last section.

Two remarks should be made. First the bias is presumably larger because our analyses neglects the noise on the analog to digital conversion of the data. Second, the maximum error used is most likely much smaller. It depends on the tuning of the analog computer. If it is optimally tuned the total error on a signal after passing through all 64 integrators is smaller than .003.

Taking these two remarks into account it still should be clear that, although we don't have exact figures, the bias on the criterium function will be positive but small.

II. Possible Interpretations of the Models

II.1. Continuous Data

The data continuousation of section (2.3) was necessary for application of data to the analog computer.
Here an economic interpretation for continuous models is given.

The data available are discrete and there are generally two possibilities with respect to the way they are published. One is as the integral of a flow variable (f.i. consumption) over the time unit used (month, year). The other possibility is as instantanuous variable (rate of intrest) measured at a point of time. However most data used here, even if they are in principle instantanuous like prices, can be treated in the same way as flow variables because they are based on series of readings between t-1 and t. The increases are based on

$$x_t = \sum_{j=1}^{N} \frac{1}{N_j} \sum_{j=0}^{N_j} (x_{jt-1+\frac{i}{N_j}} - x_{jt-1+\frac{i-1}{N_j}}) / x_{jt-1+\frac{i-1}{N_j}}$$

where M is the number of different goods in the basket for x and N_j the number of readings for goods of category j. (see CBS 1976)

Because of this most of the macro economic variables used here can be treated as flow variables. For these variables we have available $\int_{t-1}^{t} x(\theta)\, d\theta = x_t$. This can be rewritten as

$$x_t = x_{t-1} \quad (1 + \int_{t-1}^{t} \dot{x}(\theta) d\theta)$$

or

$$\Delta \log (x_t) = \log (1 + \int_{t-1}^{t} \dot{x}(\theta) d\theta)$$

In this report discrete and continuous models are studied. For discrete models the way the data we made continuous is of no intrest because the magnitude used at a point in computer time is equivalent to the value of the integral of the flow variable. For continuous models such a point in computer time is equivalent to sampling the flow or a transformation of it. We don't have data on the flow and to obtain an economically acceptable interpretation we have to make a simplifying assumption. It is assumed that the growth rate for year t is equal to $\dot{x}(\tau)$ at $t-\frac{1}{2}$. In symbols $\dot{x}(\tau) = \frac{X_t - X_{t-1}}{X_{t-1}}$ for $\tau = \frac{1}{2}$.

In case of linear growth this assumption is exact but in all other cases we introduce an error. However this assumption makes it possible to use all the functional forms given in section (II.4). In this way we have "constructed" instantanuous data for flow variables i.e. $\dot{x}(\tau)$ is an observation on $\dot{x}(t)$.

II.2. Adaptive expectations

The continuous formulation of the adaptive expectations hypothesis is

$$\text{(II.2.1)} \quad \begin{aligned} \frac{dy^*(t)}{dt} &= \beta(y(t) - y^*(t)); \ \beta \in R^+ \\ y^*(0) &= y_0^* \end{aligned}$$

In this formulation $y(t)$ and $y^*(t)$ are defined in levels.
A natural assumption is the consistency axiom which states: "... that the expectations formed at time t for the same time t must equal the actual value prevailing at that time". (Burmeister and Turnovsky [1976]). They showed that the only condition under which (1) can be justified if one discards the weak consistency axiom which implies $y^*(t) = y(t)$.

However most formulations are not in levels but in rate of changes. This means there are two possibilities with respect to the consistency axiom. The first one is with respect to the levels and the second one with respect to the rate of changes. This last possibility requires the forecast of the current rate of change to be equal to the actual rate of change. As is shown by Burmeister and Turnovsky the assumption on the rates of changes implies the assumption on the levels but not vice versa. For this reason the assumption $\dot{y}^*(t) = \dot{y}(t)$ is called the weak consistency axiom and $y^*(t) = y(t)$ the strong consistency axiom. It can further be shown that under the assumptions

$$\dot{y}^*(t) = \dot{y}(t) \text{ and } \lambda \text{ finite,}$$

the continuous adaptive expectations hypothesis

$$\text{(II.2.2)} \quad \begin{aligned} \frac{d\dot{y}^*(t)}{dt} &= \lambda(\dot{y}(t) - \dot{y}^*(t)) \\ \dot{y}^*(0) &= \dot{y}_0^* \end{aligned}$$

does make sense. The dots will be omited from now on.

The discrete version of the adaptive expectations hypothesis is, that

(II.2.3)
$$x^*_\tau - x^*_{\tau-1} = \gamma(x_t - x^*_{\tau-1}),\ \gamma \quad (0,1),\ \tau = 1,2,3,\ldots,$$
$$x^*_0 = v$$

There are many possible choices for y(t) to generate a solution which implies*)

$$y^*(\tau) = x^*_\tau, \quad \tau = 0,1,2,\ldots$$

In case of linear interpolation of x_τ the solution is

$$\lambda := -\log\ (1-\gamma)$$

$$y(t) := \frac{-\gamma}{\log(1-\gamma)}[(1-\gamma)^t x_0 + \sum_{i=0}^{[t]} (1-\gamma)^i \{x_{[t]+1-i} - x_{[t]-i}$$

$$(\log(1-\gamma)*((1-t_f)x_{[t]-i} + t_f\ x_{[t]+1-i})\}\ ;\ t \in R^+ \cap N$$

and

$$y(t) := \frac{-\gamma}{\log(1-\gamma)}\ [(1-\gamma)^t x_0 + \sum_{i=0}^{t-1} (1-\gamma)^i \{x_{t+1-i} - x_{t-i} -$$

$$(\log(1-\gamma))*x_{t-i}\}\ ;\ t \in N$$

Here [t] is the entier of t and t_f = t-[t]. What is needed is x_0, x_1, x_2,... . This y(t) is in general not continuous.

*) The solution used here is suggested by Prof. L.R.J. Westermann. For a more general treatment see Westermann [1980].

II.3. Discrete interpretations

Model (I.5.1) is formulated in continuous time. In this section it will be shown that also results for discrete models, which are mostly used in applied econometrics, can be obtained by hybrid optimization. If one samples from the implemented model the sample results can also be interpreted in discrete time. The continuity of the physical representation is now meaning less and only the values at the sample points $\tau = 1,2,...,N$ are of interest. Some of these possible interpretations are reported here.

Linear Models

With the assumption $\lambda_i = 0$, $i = 1,...,p$ model (I.5.1) reduces to

$$y(t) = \alpha_1^T v(t-\lambda)$$

$\bar{y}(t) \in R$; $\bar{\alpha}(t) \in R^{q+1}$; $v(t-\lambda) = (1, v_1(t-\lambda_1),...,v(t-\lambda_q))$; $\lambda_i \geq 0$, $i = 1,...,p$. If the least squares solution is sampled at $\tau = 1,...,N$ *) this solution is "near"**) the OLS result for

$$\text{(II.3.1)} \qquad y_\tau = \alpha_1^T x_{\tau-\lambda} + u_\tau;\ u_\tau \text{ i.i.d.}$$

y_τ, $u_\tau \in R$; $\alpha_1 \in R^{q+1}$, $x_{\tau-\lambda} = (1, x_{1\tau-\lambda_1},...,x_{q\tau-\lambda_q})^T$.

In this way also all kinds of other discrete models which can be reduced to OLS by means of a transformation can be obtained. The transformation of the data can be performed digitally and the transformed data are then applied to the hybrid computer. In this way consistent estimates for a number of "traditional" econometric models can be obtained.

*) The scaling factors are disregarded here, taking them into account would only complicate the notation.
**) This "near" was subject of the last paragraph.

Expectations models

In his book "Distributed Lags" Dhrymes [1981] describes a procedure to estimate

$$\text{(II.3.2)}\quad \begin{aligned} & y_t = \alpha_0 + \alpha_1 x^*_t + \mu_t;\ \mu_t \quad N(0,\sigma^2);\ t = 1,\ldots,T \\ & (x^*_t - x^*_{t-1}) = \lambda(x_t - x^*_t) \end{aligned}$$

by first rewritting

$$\begin{aligned} \lambda \sum_{i=0}^{\infty} (1-\lambda)^i x_{t-i} &= \lambda \sum_{i=0}^{t-1} (1-\lambda)^i x_{t-1} + \lambda(1-\lambda)^t \sum_{i=0}^{\infty} (1-\lambda)^i x_{t-i} \\ &= \lambda \sum_{i=0}^{t-1} (1-\lambda)^i x_{t-i} + \lambda(1-\lambda)^t \beta_0 \end{aligned}$$

where β_0 is the truncation reminder. Introduction of this in (II.3.2a) leads to

$$\text{(II.3.3)}\quad y_t = \alpha_0 + \beta_0\lambda(1-\lambda)^t + \alpha_1\lambda \sum_{i=0}^{t-1} (1-\lambda)^i x_{t-i} + \mu_t;\ t = 1,\ldots,T$$

If it is assumed that $x^*_0 = 0$ and $x^*_1 = x_1$ the matrix of observations can easily be formed. Because $\lambda \in (0,1)$ the term $\beta_0\lambda(1-\lambda)^t$ can be neglecked for t large.

The solution used here is slightly different. By starting at $i = -n$, n N and with the assumptions $x^*_{-n} = x_{-n}$, $x^*_{-(n+1)} = 0$ and $\beta_0\lambda(1-\lambda)^n \sim 0$ *) the relation changes into

$$\text{(II.3.4)}\quad y_t = \alpha_0 + \alpha_1\lambda \sum_{i=0}^{t-1} (1-\lambda)^i x_{t-i} + u_t;\ t = 1,\ldots,T$$

The assumption $x^*_{-(n+1)} = 0$ can easily be replaced by $x^*_{-(n+1)} = a$, which leads to $x^*_1 = x_1 + (1-\lambda)a$.
Starting from

*) If necessary the term $\beta_0\ (1-\lambda)^{t+n}$ could be introduced in the patching by using the track store.

$$\text{(II.3.5)} \quad \begin{aligned} y_t &= \alpha_0 + \sum_{j=1}^{p} \alpha_j \, x^*_{jt} + \mu_t; \ \mu_t \in N(0,\sigma^2), \ t = 1,\dots,T \\ (x^*_{jt} - x^*_{jt-1}) &= \lambda_j(x_{jt} - x^*_{jt}); \ j = 1,\dots,p. \end{aligned}$$

and with straight foreward generalisations of the assumptions made for the one variable case one can get

$$\text{(II.3.6)} \quad y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j \sum_{i=0}^{t-1} \lambda_j^i \, x_{jt-i} + \mu_t$$

The relation between the solution for x^*_{jt} and the representation used in the patching is given in section (II.2).*)

One can also use the sampled signals from the solution of the differential euqation

$$\text{(II.3.7)} \quad \begin{aligned} \frac{dx^*(t)}{dt} &= \lambda(x(t) - x^*(t)), \ \lambda \in R^+ \\ x^*(0) &= x^*_0 \end{aligned}$$

as a form of discrete expectations although the interpretation is not as clear as in the adaptive expectations case. The solution is $x^*_\tau = x^*(\tau)$; $\tau \quad N$.

Relation (II.3.5) is easily widened to

$$\text{(II.3.8)} \quad y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j x^*_{jt} + \sum_{h=1}^{q} \beta_h \, z_{ht} + \mu_t$$

Here z_{ht} can also be a delayed variable.

*) If necessary the solution of $(x^* - x_{t-1}) = (x_t - x_{t-1})$ can also be implemented on the patchboard by using the track store.

II.4. Continuous Interpretations

Linear Models

The first model is

$$(II.4.1) \qquad y(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_i(t) + \varepsilon(t)$$

with $E\{\varepsilon(t)\} = 0$, $E\{\varepsilon(t), \varepsilon(s)\} = \delta_{t\delta}\sigma^2$

The data for $\tau \in N$ are either instantaneous variables or the assumption of section (II.1) is applied.

The Phillips form

In 1954 and 1958 A.W. Phillips published two articles on a model containing relations of the form*)

$$(II.4.2) \qquad \begin{aligned} & y(t) = \alpha_0 + \sum_{j=1}^{p} \alpha_j \overset{*}{x}_j(t) + \sum_{h=1}^{q} \beta_h z_h(t) + \varepsilon(t); \; t \quad [0,T] \\ & \frac{d\overset{*}{x}_j(t)}{dt} = \lambda_j(x_j(t) - \overset{*}{x}_j(t)); \; j = 1,\ldots,p \\ & \overset{*}{x}_j(0) = x_{j0} \end{aligned}$$

It is immediate that there is a relation between this form and the form (II.3.5). The only difference is with respect to the assumption on the data. The assumption here is the same as for the previous model.

Differential Forms

The first model considered is

$$(II.4.2) \qquad dy(t) = \gamma y(t) + \beta z^T(t) + \mu(t)$$

*) The model introduced by Phillips was non stochastic.

with $\gamma \in R$ and $\beta \in R^q$ fixed coefficients, $y(t) \in R$ the endogenous variable and $z(t) \in R^q$ the vector of exogenous variables which we regard as continuous functions of time.*) The unobservable random variable u(t) is considered to satisfy $\mu(t) = \frac{d\xi(t)}{dt}$ with $\xi(t)$ an additive white noise process with uncorrelated increments.

The solution of (II.4.2) is

$$(II.4.3) \qquad y(t) = y(0)e^{\gamma t} + \int_0^t e^{\gamma(t-\theta)}\beta^T z(\theta)d\theta - \int_0^t e^{\gamma(t-\theta)}d\xi(\theta).$$

By substracting $e^{\delta y}y(t-\delta)$, with δ the time unit, and change of variable in the integrals one obtains

$$(II.4.4) \qquad y(t) = y(t-\delta)e^{\gamma\delta} + \int_0^\delta e^{\gamma\theta}\beta^T z(t-\theta)d(\theta) + \int_0^\delta e^{\gamma\theta}d\xi(t-\theta)$$

and by defining $y_\tau = y(\tau\delta)$, $\tau \in Z$, one gets the exact discrete analogue. The observations of system (II.5.2) will satisfy

$$(II.4.5) \qquad y_\tau = y_{\tau-\delta}\, e^{\gamma\delta} + \int_0^\delta e^{\gamma\theta}\, \beta^T z(\tau\delta-\theta)d\theta + \int_0^\delta e^{\gamma\theta}\, d\xi(\tau\delta-\theta)$$

The problem which remains is the treatment of the exogenous variables z(t). Because they are not simple functions of time it is necessary to approximate the second term in the righthand side of the exact discrete analoge. This can be done by approximating it by a quadratic in θ for $\theta \in [0,\delta]$ i.e. $z(\tau\delta-\theta) = a\theta^2+b\theta+c$ (see Gandolfo (1981) pg. 77-79 or Phillips (1976)).
Introduction in (II.4.5) leads to

$$(II.4.6) \qquad y_\tau = y_{\tau-1}\, e^{\gamma} + \sum_{i=0}^{2} e_i z_{\tau-i} + w_\tau$$

with

$$e_0 = [(1+\tfrac{1}{2}\gamma)c + \tfrac{1}{4}\ \beta$$
$$e_1 = [(-2+\gamma^2)c + 1 + \tfrac{1}{2}\gamma\ \beta$$
$$e_2 = [(1-\tfrac{1}{2}\gamma)c - \tfrac{1}{4}\ \beta$$
$$c = \sum_{k=0}^{\infty} \gamma^k[(k+3)!]^{-1}$$

(Gandolfo pag. 81)

*) For other assumptions on the exogenous variables see Sargan [1976].

This model can by used for estimation if $e^{\gamma\theta}$ is neglected in the error term.

An obvious disadvantage for practical calculations is that the righthand side contains y_{t-1}, z_t, z_{t-1}, z_{t-2} which may give rise to multicollinearity.

If model (II.4.4) contains flow variables we have to integrate over the unit time interval which leads to

$$\text{(II.4.7)} \quad \frac{1}{\delta}\int_{t-\delta}^{t} y(\theta)d\theta = e^{\gamma\delta}\frac{1}{\delta}\int_{t-\delta}^{t} y(\theta-\delta)d\theta + \int_{\theta}^{\delta} e^{\gamma\theta}\int_{t-\delta}^{t} \beta^{T}z(\delta-\theta)d\delta d\theta + \frac{1}{\delta}\int_{t-\delta}^{t}\int_{0}^{\delta} e^{\gamma\theta}\, d\xi(\delta-\theta)ds$$

In the literature the error term is approximated by $\int_{t-\delta}^{t}\int_{0}^{\delta} d\xi(\delta-\theta)d\delta$. (Bergström and Wymer (1976), Gandolfo (1981) pg. 108-114). It can be shown that this can be written as $(1 + .268L)\varepsilon_{\tau}$ with ε_{τ} uncorrelated. From this it is clear that for flow variables one obtains an equation with an autocorrelated error term. If we define $y_{\tau}^{0} = \frac{1}{\delta}\int_{t-\delta}^{t} y(\theta)d\theta$ and use the same approximation for z(t) as used above, we obtain

$$y_{\tau}^{0} = e^{\gamma\delta}\, y_{\tau-\delta} + \sum_{i=0}^{2} e_i\, z_{\tau-i}^{0} + (1+.268L)\varepsilon_{\tau}$$

or

$$\text{(II.4.8)} \quad y_{\tau}^{*} = e^{\gamma\delta}\, y_{\tau-\delta}^{*} + \sum_{i=0}^{2} e_i\, z_{\tau-i}^{*} + \varepsilon_{\tau} \text{ with}$$

$$x_{\tau}^{*} = x_{\tau}^{0} - .268x_{\tau-1}^{0} + (.268)^{2}x_{\tau-2}^{0} - (.268)^{3}x_{\tau-3}$$

From all this it is clear that estimation of the parameters of model (II.5.2) is rather complicated and only estimates for approximate models can be obtained. It should also be noticed that the theory stated above is exactly the same for multivariate models and is treated as such in the references mentioned. The reason for the treatment of model (II.4.7) like this is that what follows, although implemented for the univariate case, can easily be applied to multivariate models. Or stated in another way, only the computer implementation has to be adjusted not the theory.

With the use of the analog part of the hybrid computer it is possible to implement model (II.4.2) in its original form and to simulate it over the sample period. The problem is to obtain the data as continuous functions of time. This is performed by using linear interpolation between the data points available (see section I.4). Because the results of the simulation are only sampled at those points in time for which data are available one can get parameter estimates by minimizing an appropriate criterion function.
The criterion taken here is the least squares criterion $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ where $\hat{y}_i$ is obtained by integrating $\frac{dy(t)}{dt}$ over time and use this result in the righthand side of (II.5.2). As a starting point we use y_0. From this it is clear what the implications of parallel calculations are, it means that we have available at the same time the time derivative of y(t) and y(t) itself and are able to use them to define one another.

The problem which remains is the treatment of flow variables. One possible solution is already given in section (II.1) where it was assumed that the observations $y(\tau)$ can be treated as observations at a point in time. This means that the problem reduces to the problem stated above. Another possible interpretation is that the linear (or some other) interpolation is regarded as an approximation to the true function of the variable.

If we abstract from the simularity between model (II.4.7) which parameters have to be estimated and the computer model there is a third interpretation possible. Integration of the model means that we have at

τ Z

(II.4.9) $\quad y_\tau = \gamma\, y_\tau^0 + \beta^T z_\tau^0 + \xi_\tau$ *)

and simultanuous integration of $y(\tau)$ to obtain y_τ^0 means that, using the least squares criterion, estimates for γ and β can be obtained. Outside the discrete sample points the continuous machine model doesn't have an interpretation.

*) $y^0 = \frac{1}{\delta} \int_0^\delta y(\theta)d$ and $v_\tau = y(\tau)$; this last variable is ofcourse not measurable. $\xi_\tau = \int_0^\delta u(\tau - \theta)d_\theta$.

By introducing $y(t-\delta)$ in relation (II.4.2) the model changes into a mixed difference-differential equation. With respect to this kind of models there is however no theory on the estimation problem. Even from a pure mathematical point of view there are problems to be solved.

III. Applications

III.1. Practical calculations

In the case of a standard linear model statistical theory gives criteria like the t-statistic or the F-statistic to determine whether an element (or a subset of elements) of the vector of parameters is in magnitude significantly different from a certain value (mostly zero). In a number of cases mentioned before one can make probebalistic assumptions which imply that the least squares criterion used leads to estimates of the parameters with desirable and known properties. For other models mentioned before this theory is less or not developed (and it is beyond the scope of this report to do so). This implies it is not possible to distinguish between parameter values. However because of the stochastic nature of the criterion function a "natural" and minimal requirement for the elements of the vector of parameters can be derived.

Introduction of data for y_i and $\bar{y}_i$ in (I.7.11) leads to a realisation of $S(\hat{\theta})$ for a given vector of parameters $\hat{\theta}$. Repeated sampling of $\bar{y}_i$ yields a set of realisations $\bar{y}_{ij}$; $i = 1,\ldots,N$, $j = 1,\ldots,M$. Introduction in the sum of squares leads to a sequence of realisations $S(\hat{\theta})_1,\ldots,S(\hat{\theta})_M$. This sequence is used to calculate

$$\overline{\overline{S(\hat{\theta})}} = \frac{1}{M}\sum_{i=1}^{M} \overline{S(\hat{\theta})_i} \quad \text{and} \quad \hat{\sigma}_S = \frac{1}{M-1}\sum_{i=1}^{M} \left(\overline{S(\hat{\theta})_i} - \overline{\overline{S(\hat{\theta})}}\right)^{\frac{1}{2}}$$

Respectively the mean and the variance of the sequence.
The mean is used as estimate for (I.6.3).

The sensitivity of the calculated vector of parameters induced by the stochastic nature of the sum of squares can now be estimated. This is done by varying one by one the elements of the vector of parameters and calculating those values $\hat{\theta}_{iL} = \hat{\theta}_i - \delta_1\hat{\theta}_i$ and $\hat{\theta}_{iR} = \hat{\theta}_i + \delta_2\hat{\theta}_i$ for which

$$\overline{S'}(\ldots,\hat{\theta}_{iL},\ldots) = \overline{\overline{S(\hat{\theta})}} - 2\hat{\sigma}_S$$

$$\overline{S}(\ldots,\hat{\theta}_{iR},\ldots) = \overline{\overline{S(\hat{\theta})}} + 2\hat{\sigma}_S$$

The interval $(\hat{\theta}_{iL}, \hat{\theta}_{iR})$ can be used to distinghuish between parameter values. It is impossible to distinguish between the parameter values in the interval $(\hat{\theta}_{iL}, \hat{\theta}_{iR})$ and if 0 $(\hat{\theta}_{iL}, \hat{\theta}_{iR})$ the corresponding variable was rejected as explanatory variable.

Given the 'optimal' outcome of the optimization procedure one can calculate $e_i = y_i - \bar{y}_i$, $i = 1,\ldots,N.$*) Now one can obtain

$$R2SIN = 1 - \frac{e^{2}}{y^{2}} \text{ and } R2COS = \frac{\bar{y}^{2}}{y^{2}},$$

$e, y, \bar{y}$ R^N and . the Euclidian norm. R2SIN and R2COS are used as measures of correlation.

If

$$R2SIN = R2COS \text{ and } \sum_{i=1}^{N} e_i = 0$$

the result is equal to the OLS result.

*) It would be more precise to calculate e_i also M times and then calculate the mean.

III.2. Some applications

(i) Comparison with OLS.

The model estimated is the import equation from the model developed on the hybrid computer (see v. Groenendaal (1981)). It states that the imports (M) depend on total demand (VA), the relative price (PM-PA), and stocks (NA).

The results are

	VA	PM-PA	ΔNA	C	
OLS	1.265	- .351	2.677	-.323	R^2 = .986
t:	8.831	-3.565	7.950	-.357	
H.S*)	1.223	- .337	3.256	.015	R2COS = .966
LHS	1.219	- .362	3.056	-.235	R2SIN = .977
RHS	1.240	- .312	3.306	.265	

*) H.S. means hybrid simulation. (LHS, RHS) is the interval for which the object function does not change significantly.

The results for hybrid simulation are obtained by one run of the program.*) Using the t-value we see that all values obtained by hybrid simulation are in the $2\hat{\sigma}$-interval of the OLS coefficients. In table (II.2.1) the results per year are shown.

*) In a number of cases the results could be improved by a second run using the result of the first run as a starting point.

The results are

TABLE III.2.1.

YEAR	DATA	OLS	HS	1-2	1-3
1954	24.016	22.829	22.623	1.187	1.393
55	8.689	10.712	11.070	-2.032	-2.381
56	13.934	8.927	9.040	5.007	4.894
57	2.880	3.750	4.178	- .869	-1.297
58	-5.885	-4.457	-3.843	-1.428	-2.042
59	13.498	11.118	10.950	2.380	3.403
1960	16.899	19.200	19.378	-2.301	-2.479
61	7.540	6.284	7.098	1.256	.443
62	5.741	5.333	5.410	.407	.331
63	0.567	8.535	8.398	1.032	1.169
64	15.942	17.940	17.893	-1.998	-1.950
65	5.635	8.668	8.323	-3.033	-2.688
66	6.009	5.366	5.413	.643	.596
67	6.490	8.033	7.835	-1.543	-1.345
68	12.722	12.406	12.075	.317	.647
69	15.453	14.752	14.933	.702	.521
1970	15.738	13.127	12.900	2.612	2.838
71	5.990	5.315	5.938	.675	.053
72	5.224	6.101	5.798	- .877	- .573
73	12.663	11.790	12.313	.873	.351
74	- .996	- .576	- .508	- .420	- .488
75	-5.381	-6.597	-6.885	1.216	1.504
76	9.959	11.918	10.905	-1.958	- .946
77	2.525	3.421	4.105	- .896	-1.580
78	5.508	6.466	6.670	- .959	-1.624

A better example is the wage equation. Gross wage per labourer (LPM) is explained by the price of private consumption (PC), productivity increase (VA-A) and the increase in the unemployment rate ($\Delta\tilde{w}$).

	PC	(VA-A)	$\Delta\tilde{w}$	₵	
OLS	1.062	.534	-1.074	2.479	R^2 = .988
t:	(6.96)	(2.68)	(-1.35)	(2.12)	
HS	1.056	.592	-1.088	2.240	R2COS = .979
LHS	(1.029)	(.552)	(-1.728)	(2.193)	R2SIN = .976
RHS	(1.083)	(.612)	(- .928)	(2.440)	

The residuals of both results are given in figure (III.2.1). Here too the HS-results are in the $2\hat{\sigma}$-interval of the OLS results.

Although there are only two examples given here, other experiments showed that the approximations are reasonable but can be improved upon. This could be done by using another optimization routine*) or by using a stronger criterium in the existing one.

(ii) In this example it is assumed that gross wage per labourer depends on the expectation of the variables mentioned above and on the expectation for the increase in total pressure on wages (PTLN). The result is

$$\begin{array}{llllll} LPM = & 2.236 + & 1.070\, PC^* + & .497(VA-A)^* + & .309\ (\Delta PTLN)^* - & 1.106(\Delta\tilde{w})^* \\ & (2.036) & (1.016) & (.457) & (.289) & (-1.186) \\ & (2.436) & (1.123) & (.557) & (.329) & (-1.112) \end{array}$$

$$\frac{dPC^*}{dt} = \underset{\substack{(.075)\\(.072)}}{.073}(PC-PC^*) \qquad \frac{d(VA-A)^*}{dt} = \underset{\substack{(.167)\\(.164)}}{.167}\ ((VA-A)-(VA-A)^*)$$

$$\frac{d(\Delta PTLN^*)}{dt} = \underset{\substack{(.120)\\(.117)}}{.118}(\Delta PTLN-\Delta PTLN^*) \qquad \frac{d(\Delta\tilde{w})^*}{dt} = \underset{\substack{(.173)\\(.164)}}{.167}(\Delta\tilde{w}-\Delta\tilde{w}^*)$$

R2COS = .971 R2SIN = .974 $\bar{e}$ = -.014

*) For this reason the NAG library was ordered for the PDP 11/45 and could be used in the future.

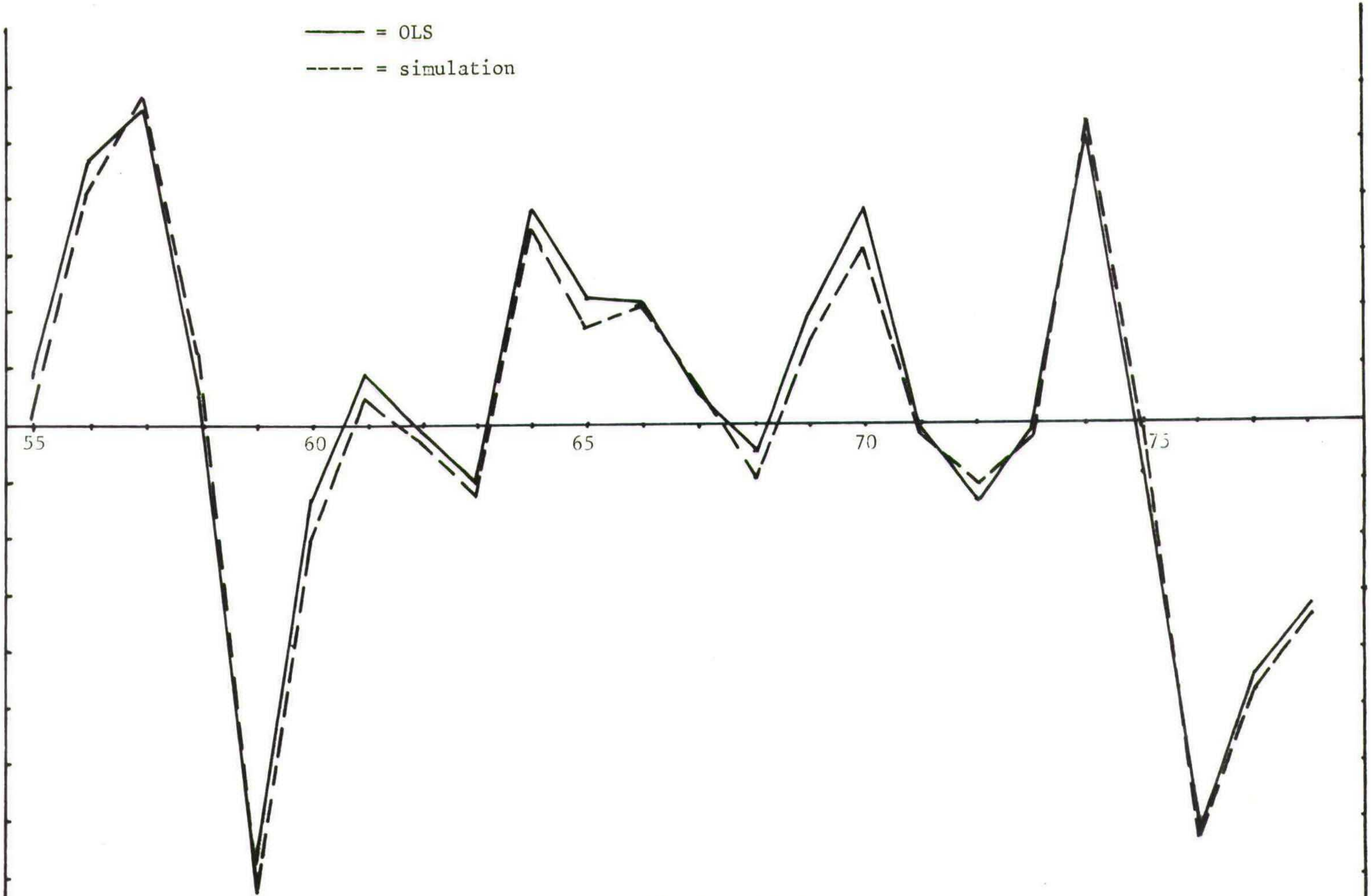


fig. III.2.1.1

The results and the residuals are given in table (III.2.2) as HS(S).

(iii) In this example a differential wage equation with the same explanatory variables as in the last example is estimated.

$$\frac{dLPM}{dt} = \underset{\substack{(-.354)\\(-.349)}}{-.351}LPM + \underset{\substack{(.368)\\(.374)}}{.375}PC^* + \underset{\substack{(.142)\\(.182)}}{.162}(VA-A)^* + \underset{\substack{(.016)\\(.025)}}{.021}(\Delta PTLN)^* - \underset{\substack{(-.458)\\(-.429)}}{.457}(\Delta\tilde{w})^* + \underset{\substack{(.939)\\(1.033)}}{.986}$$

$$\frac{dPC^*}{dt} = \underset{\substack{(.081)\\(.081)}}{.081}(PC-PC^*) \qquad \frac{d(VA-A)^*}{dt} = \underset{\substack{(.147)\\(.146)}}{.146}((VA-A)-(VA-A)^*)$$

$$\frac{d(\Delta PTLN)^*}{dt} = \underset{\substack{(.229)\\(.219)}}{.224}(\Delta PTLN-\Delta PTLN^*) \qquad \frac{d(\Delta\tilde{w})^*}{dt} = \underset{\substack{(.161)\\(.158)}}{.160}(\Delta\tilde{w}-\Delta\tilde{w}^*)$$

R2COS = .976 R2SIN = .975 $\bar{e} = -.030$

The results and the residuals are given in table (III.2.2) as HS(D).

From the table we see that, although the R2COS and the R2SIN are nearly the same, there are remarkable differences per year. However, only in one year (1962) the signs of the residuals are opposite.

TABLE III.2.2.

YEAR	DATA	HS(S)	HS(D)	1-2	1-3
1954	8.967	10.012	10.066	-1.045	-1.099
55	8.948	7.624	8.580	1.324	.368
56	8.775	5.928	6.608	2.847	2.167
57	11.775	8.774	8.846	3.001	2.929
58	4.403	5.550	4.836	-1.147	- .433
59	2.369	6.472	6.500	-4.103	-4.131

1960	8.145	8.572	8.736	- .427	- .591
61	7.236	6.278	7.026	.958	.208
62	5.872	5.656	6.140	.216	- .268
63	8.963	9.342	9.284	- .379	- .321
64	14.947	13.362	12.850	1.585	2.097
65	11.144	9.872	10.252	1.272	.892
66	10.973	9.964	9.936	1.009	1.037
67	8.767	8.376	8.428	.391	.339
68	8.872	9.366	9.058	- .494	- .186
69	13.434	12.716	12.316	.718	1.118
1970	13.224	11.360	11.482	1.834	1.742
71	13.133	12.960	12.908	.173	.225
72	12.589	13.624	13.124	-1.035	- .535
73	15.589	15.912	15.478	- .323	.110
74	15.670	13.444	13.456	2.226	2.214
75	12.812	13.628	13.350	- .816	- .538
76	10.990	14.504	14.576	-3.514	-3.586
77	8.294	10.142	10.956	-1.848	-2.662
78	7.209	8.402	9.052	-1.193	-1.843

IV. Concluding remarks

1. The relations implemented cover a wide range of models used in applied econometrics and in a number of cases the estimations obtained have desirable properties from a statistical point of view.

2. Of course a number of traditional models can more easily be estimated using the traditional package on a digital computer but with respect to models with several expectations variables as explanatory variables the program has a number of possibilities which are usually not implemented.

3. With respect to time continuous models the possibility of univariate estimation is of interest. Especially because it is possible to estimate the univariate specification directly where the methods suggested in the literature need more complicated approximations. Complicated here means that the model has to be transformed to obtain an estimable specification. A more careful comparison of the different methods is however necessary to obtain more information on the ins and outs of the method suggested here.

4. Because the hybrid computer works in continuous time and because of the calculation components it contains it is evident that its possibilities lie in the construction of time continuous models.

5. Although we have suggested some possibilities with respect to data continuation a more detailed study is necessary to give a solid foundation for practical purposes.

6. From the practical applications it is clear that although the results are reasonable, an improvement of the optimization routine, would most likely lead to better results.

7. The (improved) residuals and other output should be used to program a number of statistical tests like t-values for the coefficients, F-statistics etc. The user would then have a total package for a wide range of models.

8. In the appendix the computer program and a users "manual" is given.

Literature

1976 Burmeister E. and S.J. Turnovsky: The Specification of Adaptive Expectations in Continuous Time Dynamic Economic Models.
Econometrica, vol. 44, no. 5.

1976 Centraal Bureau voor de Statistiek: Prijsindex voor de Gezinsconsumptie I en II.
Maandstatistiek van de prijzen; pg. 250-254 en pg. 297-299.

1981 Dhrymes P.J.: Distributed Lags.
North Holland.

1974 Eykhoff P.: System Identification.
John Wiley and Sons.

1981 Gandolfo G.: Qualitative Analysis and Econometric Estimation of Continuous Time Dynamic Models.
North Holland.

1982 Groenendaal van W.: Building and Analyzing an Econometric Model with the Use of a Hybrid Computer. Discussion Paper no. 82.01, Tilburg University.

1954 Phillips A.W.: Stabilisation Policy in a Closed Economy.
Economic Journal.

1958 Phillips A.W.: Stabilisation Policy and the Time-Forms of Lagged Responses.
Economic Journal.

1976 Phillips P.C.B.: The Estimation of Linear Stochastic Differential Equations with Exogenous Variables.
In A.R. Bergström ed.: Statistical Inference in Continuous Time Economic Models.
North Holland.

1976 Sargan J.D.: Some Discrete Approximations to Continuous Time Stochastic Models.
In A.R. Bergström ed.

1979 Sylwestrowicz J.D.: Applications of Hybrid Computers in Econometrics; Part II.
Discussion Paper nr. 79.088; Tilburg University.

1980 Westermann L.R.J.: On Sequences of Difference Equations Limiting to Differential Equations.
Working Paper no. 82 (WS-8101); University of Groningen.

APPENDIX

The appendix is available on request.
It contains a description of the interactive part of the program and the output. It also contains the digital routines and the patching.

IN 1983 REEDS VERSCHENEN

126 H.H. Tigelaar
Identification of noisy linear systems with multiple arma inputs.

127 J.P.C. Kleijnen
Statistical Analysis of Steady-State Simulations: Survey of Recent Progress.

128 A.J. de Zeeuw
Two notes on Nash and Information.

129 H.L. Theuns en A.M.L. Passier-Grootjans
Toeristische ontwikkeling - voorwaarden en systematiek; een selectief literatuuroverzicht.

130 J. Plasmans en V. Somers
A Maximum Likelihood Estimation Method of a Three Market Disequilibrium Model.

131 R. van Montfort, R. Schippers, R. Heuts
Johnson S_U-transformations for parameter estimation in arma-models when data are non-gaussian.

132 J. Glombowski en M. Krüger
On the Rôle of Distribution in Different Theories of Cyclical Growth.

133 J.W.A. Vingerhoets en H.J.A. Coppens
Internationale Grondstoffenovereenkomsten.
Effecten, kosten en oligopolisten.

134 W.J. Oomens
The economic interpretation of the advertising effect of Lydia Pinkham.

135 J.P.C. Kleijnen
Regression analysis: assumptions, alternatives, applications.

136 J.P.C. Kleijnen
On the interpretation of variables.

137 G. van der Laan en A.J.J. Talman
Simplicial approximation of solutions to the nonlinear complementarity problem with lower and upper bounds.

IN 1984 REEDS VERSCHENEN

138 G.J. Cuypers, J.P.C. Kleijnen en J.W.M. van Rooyen
Testing the Mean of an Asymetric Population:
Four Procedures Evaluated

139 T. Wansbeek en A. Kapteyn
Estimation in a linear model with serially correlated errors when observations are missing

140 A. Kapteyn, S. van de Geer, H. van de Stadt, T. Wansbeek
Interdependent preferences: an econometric analysis